

Requerimientos

Lógica de repositorio de conocimiento

En este documento se establecen las características generales del nuevo sistema de lógica de repositorio de conocimiento que se desarrolla para el Pescador.

1. Definiciones

Repositorio de conocimiento (RC) Conjunto de información estructurado según las normas de la Web Semántica. El Pescador guarda dicha información en un almacén de triples, el cual constituye el *back-end* de datos de la aplicación.

Formato de datos Un modo de organización de datos.

Formato de presentación al usuario (FPU) Un formato de datos que permite la presentación de éstos directamente ante el usuario.

FPU interactivo estándar El FPU estándar que despliega el interfaz Web del Pescador. Incluye elementos interactivos como vínculos, botones, pestañas y demás controles.

FPU para edición El FPU que se emplea en el interfaz que permite al usuario modifica los datos del RC.

FPU no interactivo estándar El FPU estándar para soportes visuales no interactivos, como el papel o un archivo de procesador de palabras.

Formato intermedio Un formato de datos que no es un FPU, sino el resultado de un paso intermedio en una cadena de transformaciones de datos en el sistema.

Lógica de RC Algoritmos para la manipulación del RC y los datos que de él provienen.

Capa de lógica de RC Capa del Pescador que implementa la lógica de RC. Una de las tareas principales de esta capa es extraer los datos del RC y llevar a cabo los primeros pasos en la cadena de transformaciones para la construcción de un FPU. Otra es modificar el RC según las indicaciones que recibe del usuario a través del interfaz y controlador.

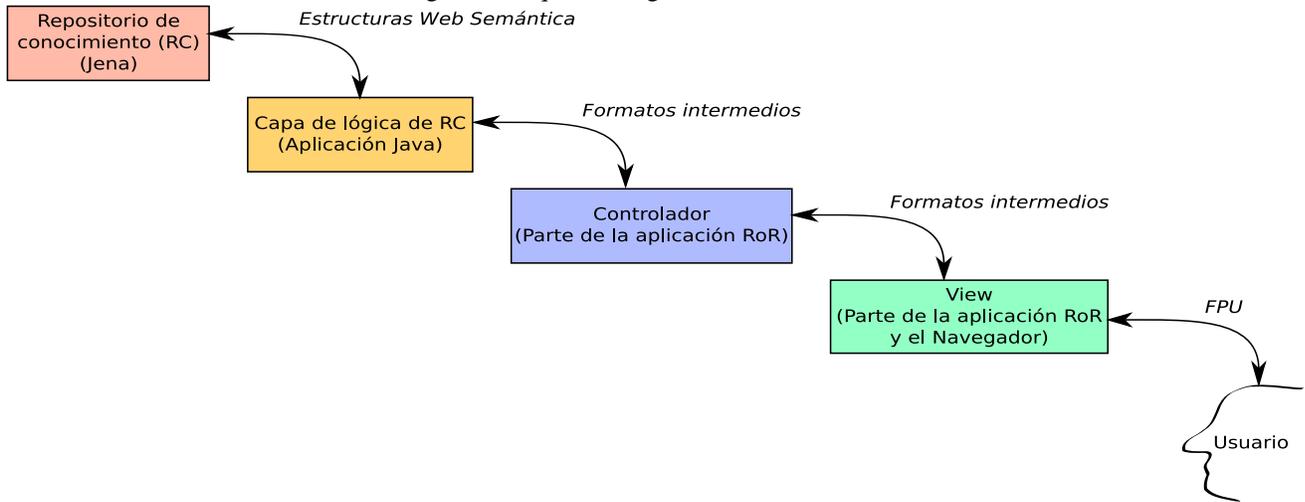
Controlador Capa del sistema, establecida según el esquema *model-view-controller*. Esta capa recibe las solicitudes del usuario y organiza los aspectos centrales de la interacción con el mismo.

Objeto del RC Recurso definido en el RC que es susceptible de ser descrito con base en los datos contenidos en el mismo RC.

2. Arquitectura general del sistema

El Pescador difunde, analiza y gestiona fondos digitales. La información acerca de los fondos se almacena en el RC. La capa de lógica RC interactúa directamente con el RC y con el controlador. En controlador interactúa con la vista, de la cual forma parte el navegador Web, que interactúa directamente con el usuario. Se trata de una arquitectura de

Figura 1: Arquitectura general del sistema



tipo *n-tier*, como se muestra en la Figura 1. En ese diagrama el texto en cursivas señala el formato de los datos al transferirse entre una capa y otra.

La capa de lógica de RC lleva a cabo los principales procesos de transformación de datos.

3. Operaciones de lógica de RC

Éstos son los cuatro tipos de operación que lleva a cabo el sistema con los datos del RC:

1. Extraer datos del RC, reorganizarlos en un FPU u otro formato, y enviarlos directamente al usuario o a otro programa.
2. Modificar los contenidos del RC con base en las indicaciones del usuario.
3. Asegurar la coherencia de las estructuras del RC.
4. Establecer, con base en las indicaciones del usuario, diversos parámetros que determinan tanto los procesos anteriores como la organización del mismo RC.

Cada uno de estos tipos generales engloba diversas categorías de operación más específicas, las cuales se describen a continuación.

Un objetivo central del diseño de la lógica de RC es permitir el reuso de mecanismos y configuraciones en la realización de diversos tipos de operación.

Extracción de datos del RC

Las operaciones que implican principalmente la extracción de datos del RC incluyen la construcción de descripciones de objetos del RC, la construcción de listas de objetos del RC y el análisis de patrones en los datos del RC.

Descripción de objetos del RC

<i>Operación</i>	<i>Detalles</i>	<i>Casos de uso</i>
Descripción larga estándar	La descripción más completa que el sistema construye de un objeto del RC por default	Mostrar registro completo
Descripción breve estándar	Descripción resumida de un objeto del RC por default	Descripciones en una lista
Descripción breve no estándar	Descripción no completa de un objeto del RC, con datos distintos a los que presenta la descripción breve estándar.	Resultados de búsqueda; Descripciones en una lista con opciones personalizadas
Descripción de etiqueta default	Descripción breve en una cadena de texto.	Componente de otras descripciones; Referencias a objetos en cadenas de texto de generación automática (p.e. en los resultados de los análisis de patrones)
Descripción para modificación de datos	Descripción de un objeto a partir de la cual se construye el FPU para edición; posiblemente revele diversas estructuras de WS que se ocultan de la mayoría de los usuarios.	Administración de un objeto o de un conjunto de objetos; Modificar descripción de un objeto o de un conjunto de objetos

Generación de listas

<i>Operación</i>	<i>Detalles</i>	<i>Casos de uso</i>
Lista de colección o grupo	Generar una lista de los objetos del RC que son miembros de una colección o grupo definidos en el mismo RC.	Ver colección; Ver grupo; Ver la lista de todas las colecciones
Lista de resultados de búsqueda	Generar una lista de objetos del RC a partir de uno o más criterios de búsqueda.	Búsqueda; Reportes automáticas
Lista de valores existentes	Generar una lista de los valores que existen en el RC para un campo determinado en un conjunto de objetos del RC.	Funciones para ayudar al usuario a determinar los criterios de una búsqueda; Funciones para generar líneas de tiempo y mapas; Agrupar los elementos de otra lista; Ir a una página determinada de una lista

Las dos primeras operaciones mencionadas—“Lista de colección o grupo” y “Lista de resultados de búsqueda”—permiten ordenar o agrupar los elementos de la lista con base en los valores de un campo determinado.

De estas operaciones la más compleja es “Lista de resultados de búsqueda”. Esta lista, además de permitir el ordenamiento y agrupación de objetos en los términos que acabamos de señalar, permite el ordenamiento por “valoración”, es decir, por la importancia de cada objeto en relación con lo buscado. Los criterios de búsqueda se definen como una serie de restricciones, cada una de las cuales devuelve un subconjunto de todos los objetos del RC. La lista de resultados final se construye por medio de la combinación de los resultados de estas restricciones, específicamente, es la intersección de todos los subconjuntos generados. Los tipos de restricción incluyen:

- restricción por búsqueda de texto libre (con mecanismos que toman en cuenta las estructuras del lenguaje natural);
- restricción por valor o rango de valores de un campo numérico determinado;
- restricción por objeto o conjunto de objetos señalados por un campo determinado;

- restricción por membresía en una colección o grupo definido en el RC;
- restricción por tipo (clase) de objeto;
- restricción por referencia temporal;
- restricción por referencia espacial; y
- restricción compuesto, cuyo resultado es la unión de los subconjuntos generados por dos o más restricciones subordinadas.

Finalmente, el mecanismo de búsqueda usa el sistema de colecciones secundarias y colecciones principales, entre otros parámetros, para determinar cuáles objetos aparecen en los resultados de búsqueda, y cuáles se constituyen como objetos “intermedios”.

Jerarquías y redes

<i>Operación</i>	<i>Detalles</i>	<i>Casos de uso</i>
Representación de jerarquía o red	Generar una representación interactiva de una jerarquía o red de objetos.	Ver tesoro o cualquier organización de datos jerárquica o en estructura de red

Análisis de conjuntos

<i>Operación</i>	<i>Detalles</i>	<i>Casos de uso</i>
Análisis de tipos	Analizar los tipos (clases) de objetos presentes en cualquier conjunto de objetos del RC.	Resumen de los contenidos de una colección, grupo o resultado de búsqueda
Análisis de relaciones	Analizar las relaciones que existen entre objetos del RC, por ejemplo, analizar las relaciones soporte-contenido y la cadena de contenidos derivados de otros.	Determinar qué contenidos son derivados (versiones) de qué otros; Determinar qué soportes plasman una representación de qué contenidos; Determinar cómo las descripciones de objetos del RC se ven impactadas por cambios en los datos referentes a otros objetos del mismo RC; Algunos pasos en la operación “Lista de resultados de búsqueda”
Análisis de descripciones posibles	Analizar las descripciones posibles para un conjunto de objetos, y determinar los elementos que tienen ellas en común.	Determinar las posibilidades de organización de un resultado de búsqueda o una lista con objeto de diferentes tipos

Otras operaciones

<i>Operación</i>	<i>Detalles</i>	<i>Casos de uso</i>
Respaldar	Generar un respaldo del RC.	Crear respaldo

Modificación de los contenidos del RC

<i>Operación</i>	<i>Detalles</i>	<i>Casos de uso</i>
Modificar la descripción de un objeto	Cambiar la información que describe a un objeto del RC.	Modificar descripción de un objeto
Modificar información administrativa de un objeto	Cambiar la información no archivística que almacena el sistema acerca de los objetos (como permisos de lectura y escritura, usuario responsable, área del RC)	Administración de un objeto
Modificar la descripción de un conjunto de objetos	Cambiar de manera simultánea la información descriptiva de un conjunto de objetos, el cual puede definirse con base en una colección, grupo o resultado de búsqueda.	Modificar conjunto de descripciones
Modificar información administrativa de un conjunto de objetos	Cambiar de manera simultánea la información no archivística de un conjunto de objetos, el cual puede definirse con base en una colección, grupo o resultado de búsqueda.	Administrar conjunto de objetos
Eliminar un objeto	Eliminar un objeto del RC y modificar apropiadamente otros objetos vinculados al primero.	Eliminar
Agregar un objeto	Generar un nuevo objeto y una estructura descriptiva del mismo, según requiere la estructura establecida para el RC.	Agregar
Dehacer / rehacer	Revertir el RC al estado que guardaba antes de una modificación determinada, o volver a realizar una modificación determinado después de haberla revertido.	Deshacer / Rehacer
Restaurar un respaldo	Restaurar un RC de una versión de respaldo.	Restaurar respaldo

La mayoría de las operaciones de modificación del RC integran las siguientes funciones: cierre de exclusión mutua (para evitar que dos usuarios intenten modificar los mismos contenidos al mismo tiempo); registrar la fecha y hora de la modificación, así como el usuario que la realizó; determinar cómo la modificación solicitada impactará en las descripciones de otros recursos que se encuentran bajo la administración de otros usuarios y organizar avisos y solicitudes de permiso, según sea apropiado; coordinar la modificación de datos bajo custodia compartida.

Asegurar la coherencia del RC

<i>Operación</i>	<i>Detalles</i>	<i>Casos de uso</i>
Revisar coherencia	Revisar el RC para verificar que los datos se encuentren estructurados de la manera establecida.	Revisar coherencia
Sugerir cambios para reparar incoherencias	En el caso de encontrar incoherencias en el RC, sugerir cambios que puedan realizarse para volver el RC a un estado coherente.	Cambios sugeridos

Establecer los parámetros de los procesos anteriores y la organización del RC

Este tipo de operación se lleva a cabo al configurar un RC y un fondo digital para su difusión con el sistema. En la presente etapa no se ofrecen funciones “amigables” para ello, sino que la configuración se realiza por medio de archivos XML/RDF. Después de implementar las demás operaciones mencionadas se definirán con mayor precisión

las operaciones de este rubro.

4. Mecanismos generales

Organización del RC y los formatos de datos

Existe una relación estrecha entre la organización de los contenidos del RC y la de los formatos de datos que genera el sistema. La capa de lógica de RC define mecanismos para establecer, a nivel de estructuras de Web Semántica, cuáles propiedades pueden y deben vincularse a cuáles objetos, y qué tipo de valores éstas pueden referir. Al mismo tiempo, definen mecanismos de transformación de los triples en formatos de datos, según las operaciones de extracción de datos arriba mencionados.

Son diversas las funciones que implementa el sistema en este nivel de mecanismos generales. Incluyen:

- El señalamiento de datos ausentes y la explicación de su ausencia
- El señalamiento de datos parciales y la explicación de por qué son parciales (por ejemplo, si no se sabe quién es el autor de una obra, pero se sabe que perteneció a una escuela determinada)
- La configuración de presentar u omitir campos determinados según si se encuentran vacíos o no
- La definición de campos repetidos y su orden por default
- La presentación de distintos valores y nombres de campos según el idioma del usuario
- En los campos de texto, la incrustación de hipervínculos hacia otros objetos del RC
- La asignación de elementos descriptivos a objetos de forma individual o por conjunto
- La definición de áreas del RC, y al interior de éstas, colecciones
- La definición de permisos de lectura y escritura, así como usuario(s) responsable(s), para áreas del RC o recursos específicos
- La definición de algoritmos complejos específicos para tipos de objeto determinados (“hooks”)
- La definición de tipos de dato (texto, imagen, datos temporales, datos espaciales, etc.) y la vinculación de éstos con determinados valores “literales” WS o con estructuras complejas de triples WS.

Cadenas de transformaciones de datos

Todos los FPU tienen algunos elementos en común, entre ellos: campos (con o sin etiqueta), secciones, subsecciones, tipos de contenido (texto, imagen, sonido, mapa, línea de tiempo, video), títulos de la descripción, de la sección o de la subsección. Se emplean formatos de datos intermedios (en lugar de crear los FPU directamente de los datos del RC) para poder abstraer estos elementos en común y los algoritmos comunes que se requieren para generarlos. En tanto, los mecanismos para generar cualquier FPU son esencialmente los mismos desde la perspectiva de la capa de lógica de RC. Son la vista y el controlador los que, a partir de los formatos intermedios de datos, agregan los elementos que distinguen un FPU determinado de los demás. (Por ejemplo, esas capas agregan pestañas, botones y demás controles para el FPU estándar interactivo, agregan saltos de página en un FPU no interactivo para impresión en papel, etc.)